

Weighted Random Popular Matchings

TOSHIYA ITOH

titoth@dac.gsic.titech.ac.jp

Global Sci. Inform. and Comput. Center
Tokyo Institute of Technology
Meguro-ku, Tokyo 152-8550, Japan

OSAMU WATANABE

watanabe@is.titech.ac.jp

Dept. of Math. and Comput. Sys.
Tokyo Institute of Technology
Meguro-ku, Tokyo 152-8552, Japan

Abstract: For a set A of n applicants and a set I of m items, we consider a problem of computing a matching of applicants to items, i.e., a function \mathcal{M} mapping A to I ; here we assume that each applicant $x \in A$ provides a *preference list* on items in I . We say that an applicant $x \in A$ *prefers* an item p than an item q if p is located at a higher position than q in its preference list, and we say that x *prefers* a matching \mathcal{M} over a matching \mathcal{M}' if x prefers $\mathcal{M}(x)$ over $\mathcal{M}'(x)$. For a given matching problem A, I , and preference lists, we say that \mathcal{M} is *more popular* than \mathcal{M}' if the number of applicants preferring \mathcal{M} over \mathcal{M}' is larger than that of applicants preferring \mathcal{M}' over \mathcal{M} , and \mathcal{M} is called a *popular matching* if there is no other matching that is more popular than \mathcal{M} . Here we consider the situation that A is partitioned into A_1, A_2, \dots, A_k , and that each A_i is assigned a weight $w_i > 0$ such that $w_1 > w_2 > \dots > w_k > 0$. For such a matching problem, we say that \mathcal{M} is *more popular* than \mathcal{M}' if the total weight of applicants preferring \mathcal{M} over \mathcal{M}' is larger than that of applicants preferring \mathcal{M}' over \mathcal{M} , and we call \mathcal{M} a *k-weighted popular matching* if there is no other matching that is more popular than \mathcal{M} . Mahdian [In Proc. of the 7th ACM Conference on Electronic Commerce, 2006] showed that if $m > 1.42n$, then a random instance of the (nonweighted) matching problem has a popular matching with high probability. In this paper, we analyze the 2-weighted matching problem, and we show that (lower bound) if $m/n^{4/3} = o(1)$, then a random instance of the 2-weighted matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching with probability $o(1)$; and (upper bound) if $n^{4/3}/m = o(1)$, then a random instance of the 2-weighted matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching with probability $1 - o(1)$.

Key Words: Random Popular Matchings, Weighted Popular Matchings, Well-Formed Matchings.

1 Introduction

For a set A of n applicants and a set I of m items, we consider the problem of computing a certain matching of applicants to items, i.e., a function \mathcal{M} mapping A to I . Here we assume that each applicant $x \in A$ provides its *preference list* defined on a subset $J_x \subseteq I$. A preference list $\vec{\ell}_x$ of each applicant x may contain ties among the items and it ranks subsets J_x^h 's of J_x ; that is, J_x is partitioned into $J_x^1, J_x^2, \dots, J_x^d$, where J_x^h is a set of the h^{th} preferred items. We say that an applicant x *prefers* $p \in J_x$ than $q \in J_x$ if $p \in J_x^i$ and $q \in J_x^h$ for $i < h$. For any matchings \mathcal{M} and \mathcal{M}' , we say that an applicant x *prefers* \mathcal{M} over \mathcal{M}' if the applicant x prefers $\mathcal{M}(x)$ over $\mathcal{M}'(x)$, and we say that \mathcal{M} is *more popular* than \mathcal{M}' if the total number of applicants preferring \mathcal{M} over \mathcal{M}' is larger than that of applicants preferring \mathcal{M}' over \mathcal{M} . \mathcal{M} is called a *popular matching* [6] if there is no other matching that is more popular than \mathcal{M} . The *popular matching problem* is to compute this popular matching for given A, I , and preference lists. This problem has applications in the real world, e.g., mail-based DVD rental systems such as NetFlix [1].

Here we consider the (general) situation that the set A of applicants is partitioned into several categories A_1, A_2, \dots, A_k , and that each category A_i is assigned a weight $w_i > 0$ such that $w_1 > w_2 > \dots > w_k$. This setting can be regarded as a case where the applicants in A_1 are platinum members, the applicants in A_2 are gold members, the applicants in A_3 are silver members, the applicants in A_4 are regular members, etc. In a way similar to the above, we define the *k-weighted popular matching problem* [8], where the goal is to compute a popular matching \mathcal{M} in the sense that for any other matching \mathcal{M}' , the total weight of applicants preferring \mathcal{M} is larger than that of applicants preferring \mathcal{M}' . Notice that the original popular

matching problem, which we will call the *single category* popular matching problem, is the 1-weighted popular matching problem.

We say that a preference list $\vec{\ell}_x$ of an applicant x is *complete* if $J_x = I$, that is, x shows its preferences on all items, and a k -weighted popular matching problem $(A, I, \{\vec{\ell}_x\}_{x \in A})$ is called *complete* if $\vec{\ell}_x$ is complete for every applicant $x \in A$. We also say that a preference list $\vec{\ell}_x$ of an applicant x is *strict* if $|J_x^h| = 1$ for each h , that is, x prefers each item in J_x differently, and a k -weighted popular matching problem is called *strict* if $\vec{\ell}_x$ is strict for every applicant $x \in A$.

1.1 Known Results

For the strict single category popular matching problem, Abraham, et al. [2] presented a deterministic $O(n + m)$ time algorithm that outputs a popular matching if it exists; they also showed, for the single category popular matching problem with ties, a deterministic $O(\sqrt{nm})$ time algorithm. To derive these algorithms, Abraham, et al. introduced the notions of f -items (the first items) and s -items (the second items), and characterized popular matchings by f -items and s -items. Mestre [8] generalized those results to the k -weighted popular matching problem, and he showed a deterministic $O(n + m)$ time algorithm for the strict case, where it outputs a k -weighted popular matching if any, and a deterministic $O(\min(k\sqrt{n}, n)m)$ time algorithm for the case with ties.

In general, some instances of the complete and strict single category popular matching problem do not have a popular matching. Answering to a question of when a random instance of the complete and strict single category popular matching problem has a popular matching, Mahdian [7] showed that if $m > 1.42n$, then a random instance of the popular matching problem has a popular matching with probability $1 - o(1)$; he also showed that if $m < 1.42n$, then a random instance of the popular matching problem has a popular matching with probability $o(1)$.

1.2 Main Results

In this paper, we consider the complete and strict 2-weighted popular matching problem, and investigate when a random instance of the complete and strict 2-weighted popular matching problem has a 2-weighted popular matching. Our results are summarized as follows.

Theorem 4.1: If $m/n^{4/3} = o(1)$, then a random instance of the complete and strict 2-weighted popular matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching with probability $o(1)$.

Theorem 5.1: If $n^{4/3}/m = o(1)$, then a random instance of the complete and strict 2-weighted popular matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching with probability $1 - o(1)$.

For an instance of the single category popular matching problem, it suffices to consider only a set F of f -items and a set S of s -items [7]. For an instance of the 2-weighted popular matching problem, however, we need to *separately* consider f_1 -items, s_1 -items, f_2 -items, and s_2 -items; let F_1 , S_1 , F_2 , and S_2 denote these item sets. Some careful analysis is necessary, in particular, because in general, we may have the situation $S_1 \cap F_2 \neq \emptyset$, which makes our probabilistic analysis much harder than (and quite different from) the single category case.

2 Preliminaries

In the rest of this paper, we consider the complete and strict 2-weighted popular matching problem. Let A be the set of n applicants and I be the set of m items. We assume that A is partitioned into A_1 and A_2 , and we refer to A_1 (resp. A_2) as the first (resp. the second) category. For any constant $0 < \delta < 1$, we also assume that $|A_1| = \delta|A| = \delta n$ and $|A_2| = (1 - \delta)|A| = (1 - \delta)n$. Let $w_1 > w_2 > 0$ be weights of the first category A_1 and the second category A_2 , respectively.

We define f -items and s -items [2, 8] as follows: For each applicant $x \in A_1$, let $f_1(x)$ be the most preferred item in its preference list $\vec{\ell}_x$, and we call it an f_1 -item of x . We use F_1 to denote the set of all f_1 -items of applicants $x \in A_1$. For each applicant $x \in A_1$, let $s_1(x)$ be the most preferred item in its preference list $\vec{\ell}_x$ that is not in F_1 , and we use S_1 to denote the set of all s_1 -items of applicants $x \in A_1$. Similarly, for each applicant $y \in A_2$, let $f_2(y)$ and $s_2(y)$ be the most preferred item in its preference list $\vec{\ell}_y$ that is not in F_1 and not in $F_1 \cup F_2$, respectively, where we use F_2 and S_2 to denote the set of all f_2 -items and s_2 -items, respectively. From this definition, we have that $F_1 \cap S_1 = \emptyset$, $F_1 \cap F_2 = \emptyset$, and $F_2 \cap S_2 = \emptyset$; on the other hand, we may have that $S_1 \cap F_2 \neq \emptyset$ or $S_1 \cap S_2 \neq \emptyset$.

For characterizing the existence of k -weighted popular matching, Mestre [8] defined the notion of “well-formed matching,” which generalizes well-formed matching for the single category popular matching problem [2]. We recall this characterization here. Below we consider any instance $(A, I, \{\vec{\ell}_x\}_{x \in A})$ of the strict (not necessarily complete) 2-weighted popular matching problem.

Definition 2.1 *A matching \mathcal{M} is well-formed if by \mathcal{M} (1) each $x \in A_1$ is matched to $f_1(x)$ or $s_1(x)$; (2) each $y \in A_2$ is matched to $f_2(y)$ or $s_2(y)$; (3) each $p \in F_1$ is matched to some $x \in A_1$ such that $p = f_1(x)$; and (4) each $q \in F_2$ is matched to some $y \in A_2$ such that $q = f_2(y)$.*

Mestre [8] showed that the existence of a 2-weighted popular matching is almost equivalent to that of a well-formed matching. Precisely, he proved the following characterization.

Proposition 2.1 ([8]) *Let $(A, I, \{\vec{\ell}_x\}_{x \in A})$ be an instance of the strict 2-weighted popular matching problem. Any 2-weighted popular matching of $(A, I, \{\vec{\ell}_x\}_{x \in A})$ is a well-formed matching, and if $w_1 \geq 2w_2$, then any well-formed matching of $(A, I, \{\vec{\ell}_x\}_{x \in A})$ is a 2-weighted popular matching.*

Consider an instance $(A, I, \{\vec{\ell}_x\}_{x \in A})$ of the strict (not necessarily complete) 2-weighted popular matching problem with weights $w_1 \geq 2w_2$. As shown above, the existence of a 2-weighted popular matching is characterized by that of a well-formed matching, which is determined by the structure of f_1 -, f_2 -, s_1 -, and s_2 -items. Here we introduce a graph $G = (V, E)$ for investigating this structure, and in the following discussion, we will mainly use this graph. The graph $G = (V, E)$ is defined by a set $V = F_1 \cup S_1 \cup F_2 \cup S_2$ of vertices, and the following set E of edges.

$$E = \{(f_1(x), s_1(x)) : x \in A_1\} \cup \{(f_2(y), s_2(y)) : y \in A_2\}.$$

We use E_1 and E_2 to denote the sets of edges defined for applicants in A_1 and A_2 , respectively, i.e., the former and the latter sets of the above. In the following, the graph $G = (V, E)$ defined above is called an *fs-relation graph* for $(A, I, \{\vec{\ell}_x\}_{x \in A})$. Note that this fs-relation graph $G = (V, E)$ consists of $M = |V| \leq m$ vertices and $n = |A|$ edges. If $e_1 \in E_1$ and $e_2 \in E_2$ are incident to the same vertex $p \in V$, then we have either $p \in S_1 \cap F_2$ or $p \in S_1 \cap S_2$. This situation makes the analysis of the 2-weighted popular matching problem harder than and different from the one for the single category case.

We now characterize the existence of a well-formed matching as follows.

Lemma 2.1 *An instance $(A, I, \{\vec{\ell}_x\}_{x \in A})$ of the strict 2-weighted popular matching problem has a well-formed matching iff its fs-relation graph $G = (V, E)$ has an orientation \mathcal{O} on edges such that (a) each $p \in V$ has at most one incoming edge in $E_1 \cup E_2$; (b) each $p \in F_1$ has one incoming edge in E_1 ; and (c) each $q \in F_2$ has one incoming edge in E_2 .*

Proof: Consider any instance $(A, I, \{\vec{\ell}_x\}_{x \in A})$ of the strict 2-weighted popular matching problem, where $A = A_1 \cup A_2$, and let $G = (V, E)$ be its fs-relation graph.

First assume that this instance has a well-formed matching \mathcal{M} . Define an orientation \mathcal{O} on edges of the graph $G = (V, E)$ as follows: For each applicant $a \in A_i$, orient an edge $e_a = (f_i(a), s_i(a)) \in E_i$ toward $\mathcal{M}(a)$. Since \mathcal{M} is a matching between A and I , we have that each $p \in V$ has at most one incoming edge. From the condition (3) of Definition 2.1, it follows that each $p \in F_1$ has one incoming edge in E_1 , and from

the condition (4) of Definition 2.1, it follows that each $q \in F_2$ has one incoming edge in E_2 . Thus the orientation \mathcal{O} on edges of $G = (V, E)$ satisfies the conditions (a), (b), and (c).

Assume that the graph $G = (V, E)$ has an orientation \mathcal{O} on edges satisfying the conditions (a), (b), and (c). Then we define a matching \mathcal{M} as follows: For each $x \in A_1$, its f_1 -item $f_1(x)$ (resp. s_1 -item $s_1(x)$) is matched to x if \mathcal{O} orients the edge $e_x = (f_1(x), s_1(x)) \in E_1$ by $f_1(x) \leftarrow s_1(x)$ (resp. $f_1(x) \rightarrow s_1(x)$), and for each $y \in A_2$, its f_2 -item $f_2(y)$ (resp. s_2 -item $s_2(y)$) is matched to y if \mathcal{O} orients the edge $e_y = (f_2(y), s_2(y)) \in E_2$ by $f_2(y) \leftarrow s_2(y)$ (resp. $f_2(y) \rightarrow s_2(y)$). From the condition (a) of the orientation \mathcal{O} , it is immediate to see that \mathcal{M} is a matching for $(A, I, \{\vec{\ell}_x\}_{x \in A})$. From the definition of the graph $G = (V, E)$, we have that \mathcal{M} satisfies the conditions (1) and (2) of Definition 2.1. The condition (b) of the orientation \mathcal{O} implies that each $p \in F_1$ is matched to $x \in A_1$ by \mathcal{M} , where $f_1(x) = p$, and the condition (c) of the orientation \mathcal{O} guarantees that each $q \in F_2$ is matched to $y \in A_2$ by \mathcal{M} , where $f_2(y) = q$. Thus the matching \mathcal{M} for $(A, I, \{\vec{\ell}_x\}_{x \in A})$ satisfies the conditions (1), (2), (3), and (4) of Definition 2.1. ■

3 Characterization for the 2-Weighted Popular Matching Problem

In this section, we present necessary and sufficient conditions for an instance of the strict 2-weighted popular matching problem to have a 2-weighted popular matching. For an instance $(A, I, \{\vec{\ell}_x\}_{x \in A})$ of the strict 2-weighted popular matching problem, let $G = (V, E)$ be its fs-relation graph, and consider the subgraphs G_1 , G_2 , and G_3 of the graph $G = (V, E)$ as in Figure 1.

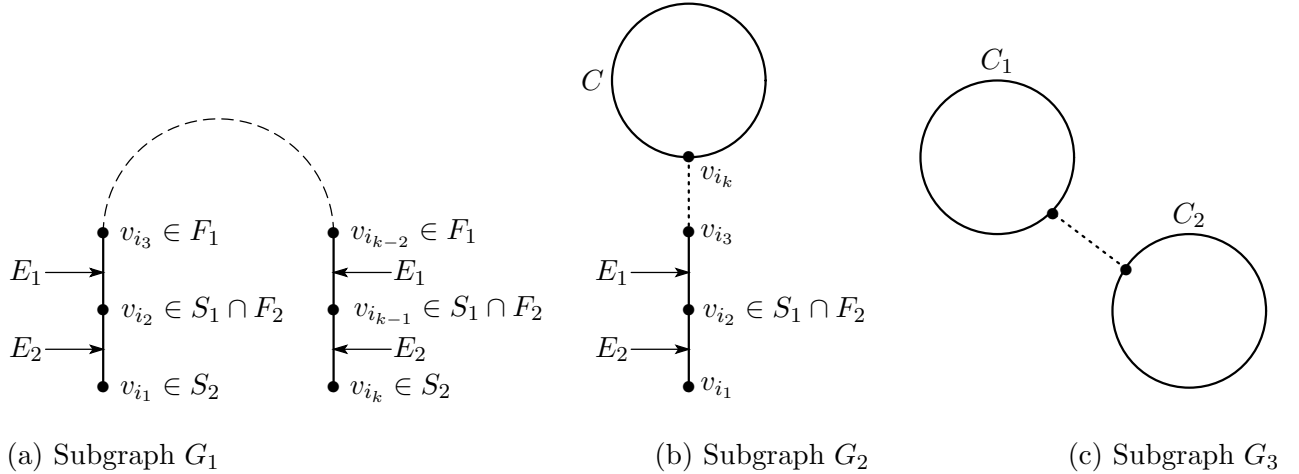


Figure 1: (a) a path $P = v_{i1}, v_{i2}, \dots, v_{ik}$ that has vertices $v_{i2}, v_{ik-1} \in S_1 \cap F_2$ such that $(v_{i2}, v_{i3}) \in E_1$ and $(v_{ik-2}, v_{ik-1}) \in E_1$; (b) a cycle C and a path $P = v_{i1}, v_{i2}, \dots, v_{ik}$ incident to C at v_{ik} that has a vertex $v_{i2} \in S_1 \cap F_2$ such that $(v_{i2}, v_{i3}) \in E_1$; (c) a connected component including cycles C_1 and C_2 .

Theorem 3.1 *An instance $(A, I, \{\vec{\ell}_x\}_{x \in A})$ of the strict 2-weighted popular matching problem has a well-formed matching iff its fs-relation graph $G = (V, E)$ contains none of the subgraphs G_1 , G_2 , nor G_3 in Figure 1.*

Proof: Assume that the graph $G = (V, E)$ contains one of the subgraphs G_1 , G_2 , and G_3 in Figure 1. For the case where the graph G contains the subgraph G_1 , if the edge $(v_{i2}, v_{i3}) \in E_1$ is oriented by $v_{i2} \leftarrow v_{i3}$, then the edge $(v_{i1}, v_{i2}) \in E_2$ is oriented by $v_{i1} \leftarrow v_{i2}$ to satisfy the condition (a) of Lemma 2.1. However, this does not meet the condition (c) of Lemma 2.1, since the vertex $v_{i2} \in S_1 \cap F_2 \subseteq F_2$ has no incoming edges in E_2 . So the edge $(v_{i2}, v_{i3}) \in E_1$ must be oriented by $v_{i2} \rightarrow v_{i3}$. It is also the case for the edge $(v_{ik-2}, v_{ik-1}) \in E_1$, that is, $(v_{ik-2}, v_{ik-1}) \in E_1$ must be oriented by $v_{ik-1} \rightarrow v_{ik-2}$. These facts imply that

there exists $2 < j < k - 1$ such that the vertex $v_{i_j} \in V$ has at least two incoming edges, which violates the condition (a) of Lemma 2.1. Thus if the graph G contains the subgraph G_1 , then the instance does not have a well-formed matching. Similarly we can show that if the graph G contains the subgraph G_2 , then the instance does not have a well-formed matching. The case where the graph G contains the subgraph G_3 can be argued in a way similar to the proof by Mahdian [7, Lemma 2].

Assume that the graph $G = (V, E)$ does not contain any of the subgraph G_1 , G_2 , or G_3 and let $\{C_i\}_{i \geq 1}$ be the set of cycles in G . We first orient cycles $\{C_i\}_{i \geq 1}$. Since the graph G does not contain the subgraph G_1 , we can orient each cycle C_i in one of the clockwise and counterclockwise orientations to meet the conditions (a), (b), and (c) of Lemma 2.1. From the assumption that the graph G does not contain the subgraph G_3 , the remaining edges can be categorized as follows: $E_{\text{tree}}^{\text{cyc}}$ = the set of edges in subtrees of G that are incident to some cycle $C \in \{C_i\}_{i \geq 1}$, and E_{tree} = the set of edges in subtrees of G that are not incident to any cycle $C \in \{C_i\}_{i \geq 1}$. Since the graph G does not contain the subgraphs G_1 and G_2 , we can orient edges in $E_{\text{tree}}^{\text{cyc}}$ away from the cycles to meet the conditions (a), (b), and (c) of Lemma 2.1. Notice that edges in E_{tree} form subtrees of G . For each such T , let E_T^2 be the set of edges (v, u) that is assigned to some applicant in A_2 and $u \in S_1 \cap F_2$. For each edge $e = (v, u) \in E_T^2$, we first orient the edge e by $v \rightarrow u$ and then the remaining edges in E_T^2 are oriented away from each $u \in S_1 \cap F_2$. By the assumption that the graph G does not contain the subgraph G_1 , such an orientation meets the conditions (a), (b), and (c) of Lemma 2.1 for each $v \in T$. \blacksquare

From Proposition 2.1 and Theorem 3.1, we immediately have the following corollary:

Corollary 3.1 *Any instance $(A, I, \{\vec{\ell}_x\}_{x \in A})$ of the strict 2-weighted popular matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching iff its fs-relation graph $G = (V, E)$ contains none of the subgraphs G_1 , G_2 , nor G_3 in Figure 1.*

Let us consider a random instance of the complete and strict 2-weighted popular matching problem. Roughly speaking, a natural uniform distribution is considered here. That is, given a set $A = A_1 \cup A_2$ of n applicants and a set I of m items, and we consider an instance obtained by defining a random preference list $\vec{\ell}_x$ for each applicant $x \in A$, which is a permutation on I that is chosen independently and uniformly at random. But as discussed above for the 2-weighted case, the situation is completely determined by the corresponding fs-relation graph that depends only on the first and second items of applicants. Thus, instead of considering a random instance of the problem, we simply define the first and second items as follows, and discuss with the fs-relation graph $G = (V, E)$ obtained defined by f_1 -, s_1 -, f_2 -, and s_2 -items.

- (1) For each $x \in A_1$, assign an item $p \in I$ as a f_1 -item $f_1(x)$ independently and uniformly at random, and let F_1 be the set of all f_1 -items;
- (2) For each $x \in A_1$, assign an item $p \in I - F_1$ as a s_1 -item $s_1(x)$ independently and uniformly at random, and let S_1 be the set of all s_1 -items;
- (3) For each $x \in A_2$, assign an item $p \in I - F_1$ as a f_2 -item $f_2(x)$ independently and uniformly at random, and let F_2 be the set of all f_2 -items; and
- (4) For each $x \in A_2$, assign an item $p \in I - (F_1 \cup F_2)$ as a s_2 -item $s_2(x)$ independently and uniformly at random, and let S_2 be the set of all s_2 -items.

It is easy to see that this choice of first and second items is the same as defining first and second items from a random instance of the complete and strict 2-weighted popular matching problem.

4 Lower Bounds for the 2-Weighted Popular Matching Problem

Let n be the number of applicants and m be the number of items. Assume that m is large enough so that $m - n \geq m/c$ for some constant $c > 1$, i.e., $m \geq cn/(c - 1)$. For any constant $0 < \delta < 1$, let $n_1 = \delta n$

and $n_2 = (1 - \delta)n$ be the numbers of applicants in A_1 and A_2 , respectively. In this section, we show a lower bound for m such that a random instance of the complete and strict 2-weighted popular matching problem has a 2-weighted popular matching with low probability.

Theorem 4.1 *If $m/n^{4/3} = o(1)$, then a random instance of the complete and strict 2-weighted popular matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching with probability $o(1)$.*

Proof: Consider a random fs-relation graph $G = (V, E)$. As shown in Corollary 3.1, it suffices to prove that $G = (V, E)$ contains one of the graphs G_1 , G_2 , and G_3 of Figure 1 with high probability. But here we focus on one simple such graph, namely, G'_1 given Figure 2, and in the following, we argue that the probability that $G = (V, E)$ contains G'_1 is high if $m/n^{4/3} = o(1)$.

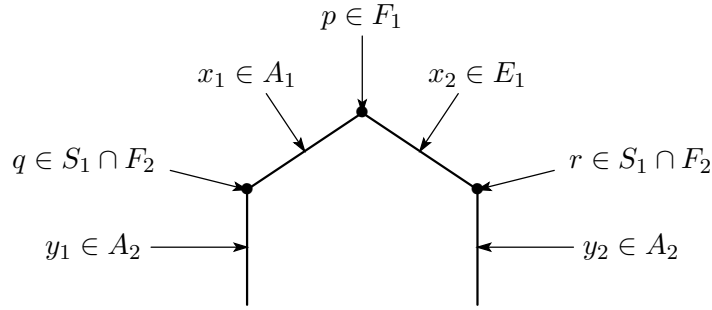


Figure 2: The Simplest “Bad” Subgraphs G'_1

Let F_1 and F_2 be the sets of the first items, S_1 and S_2 be the sets of the second items, respectively, for applicants in A_1 and A_2 . By the definitions of F_1 , F_2 , S_1 , and S_2 , we have that $F_1 \cap S_1 = \emptyset$, $F_1 \cap F_2 = \emptyset$, $F_1 \cap S_2 = \emptyset$, and $F_2 \cap S_2 = \emptyset$. On the other hand, we may have that $S_1 \cap F_2 \neq \emptyset$ or $S_1 \cap S_2 \neq \emptyset$. Let $R_1 = I - F_1$ and $R_2 = R_1 - F_2 = I - (F_1 \cup F_2)$. It is obvious that $1 \leq |F_1| \leq \delta n$ and $1 \leq |F_2| \leq (1 - \delta)n$, which implies that $m - \delta n \leq |R_1| \leq m$ and $m - n \leq |R_2| \leq m$.

For any pair of $x_1, x_2 \in A_1$ such that $x_1 < x_2$ and any pair of $y_1, y_2 \in A_2$ such that $y_1 \neq y_2$, we simply use \vec{v} to denote (x_1, x_2, y_1, y_2) , and T to denote the set of all such \vec{v} 's. Since $n_1 = \delta n = |A_1|$ and $n_2 = (1 - \delta)n = |A_2|$, we have that for sufficiently large n ,

$$|T| = \binom{n_1}{2} n_2 (n_2 - 1) \geq \frac{\delta^2 (1 - \delta)^2}{3} n^4. \quad (1)$$

For each $\vec{v} = (x_1, x_2, y_1, y_2) \in T$, define a random variable $Z_{\vec{v}}$ to be $Z_{\vec{v}} = 1$ if x_1, x_2, y_1 , and y_2 form the bad subgraph G'_1 in Figure 2; $Z_{\vec{v}} = 0$ otherwise. Let $Z = \sum_{\vec{v} \in T} Z_{\vec{v}}$. Then from Chebyshev's Inequality [9, Theorem 3.3], it follows that

$$\begin{aligned} \Pr[Z = 0] &\leq \Pr[|Z - \mathbf{E}[Z]| \geq \mathbf{E}[Z]] \\ &= \Pr\left[|Z - \mathbf{E}[Z]| \geq \frac{\mathbf{E}[Z]}{\sigma_Z} \sigma_Z\right] \leq \frac{\sigma_Z^2}{\mathbf{E}^2[Z]} = \frac{\mathbf{Var}[Z]}{\mathbf{E}^2[Z]}. \end{aligned} \quad (2)$$

To derive the lower bound for $\Pr[Z > 0]$, we estimate the upper bound for $\mathbf{Var}[Z]/\mathbf{E}^2[Z]$. We first consider $\mathbf{E}[Z]$. For each $\vec{v} \in T$, it is easy to see that

$$\begin{aligned} \Pr[Z_{\vec{v}} = 1] &\geq \frac{1}{m} \cdot \left(\frac{1}{m}\right)^2 = \frac{1}{m^3}; \\ \Pr[Z_{\vec{v}} = 1] &\leq \frac{1}{m} \cdot \left(\frac{1}{m - n_1}\right)^2 \leq \frac{1}{m} \cdot \left(\frac{1}{m - n}\right)^2 = \frac{c^2}{m^3}, \end{aligned} \quad (3)$$

where Inequality (3) follows from the assumption that $m - n_1 \geq m - n \geq m/c$ for some constant $c > 1$. Thus from the estimations for $\Pr[Z_{\vec{v}} = 1]$, it follows that

$$\mathbf{E}[Z] = \mathbf{E}\left[\sum_{\vec{v} \in T} Z_{\vec{v}}\right] = \sum_{\vec{v} \in T} \mathbf{E}[Z_{\vec{v}}] = \sum_{\vec{v} \in T} \Pr[Z_{\vec{v}} = 1] \geq \frac{|T|}{m^3}; \quad (4)$$

$$\mathbf{E}[Z] = \mathbf{E}\left[\sum_{\vec{v} \in T} Z_{\vec{v}}\right] = \sum_{\vec{v} \in T} \mathbf{E}[Z_{\vec{v}}] = \sum_{\vec{v} \in T} \Pr[Z_{\vec{v}} = 1] \leq \frac{c^2|T|}{m^3}. \quad (5)$$

We then consider $\mathbf{Var}[Z]$. From the definition of $\mathbf{Var}[Z]$, it follows that

$$\begin{aligned} \mathbf{Var}[Z] &= \mathbf{E}\left[\left(\sum_{\vec{v} \in T} Z_{\vec{v}}\right)^2\right] - \left(\mathbf{E}\left[\sum_{\vec{v} \in T} Z_{\vec{v}}\right]\right)^2 \\ &= \mathbf{E}\left[\sum_{\vec{v} \in T} Z_{\vec{v}}^2 + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T - \{\vec{v}\}} Z_{\vec{v}} Z_{\vec{w}}\right] - \left(\mathbf{E}\left[\sum_{\vec{v} \in T} Z_{\vec{v}}\right]\right)^2 \\ &= \mathbf{E}\left[\sum_{\vec{v} \in T} Z_{\vec{v}}\right] - \left(\mathbf{E}\left[\sum_{\vec{v} \in T} Z_{\vec{v}}\right]\right)^2 + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T - \{\vec{v}\}} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \\ &= \mathbf{E}[Z] - \mathbf{E}^2[Z] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T - \{\vec{v}\}} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}]. \end{aligned} \quad (6)$$

In the following, we estimate the last term of Equality (6). For each $\vec{v} = (x_1, x_2, y_1, y_2) \in T$ and each $0 \leq h \leq 2$, we say that $\vec{w} = (x'_1, x'_2, y'_1, y'_2) \in T - \{\vec{v}\}$ is h -common to \vec{v} if $|\{x_1, x_2\} \cap \{x'_1, x'_2\}| = h$. For any $\vec{w} = (x'_1, x'_2, y'_1, y'_2) \in T$ that is 2-common to \vec{v} , we have that $x_1 = x'_1$ and $x_2 = x'_2$, because if $x_1 = x'_2$ and $x_2 = x'_1$, then $x_1 = x'_2 > x'_1 = x_2$, which contradicts the assumption that $x_1 < x_2$. For each $\vec{v} \in T$, we use $T_2(\vec{v})$ to denote the set of $\vec{w} \in T - \{\vec{v}\}$ that is 2-common to \vec{v} ; $T_1(\vec{v})$ to denote the set of $\vec{w} \in T - \{\vec{v}\}$ that is 1-common to \vec{v} ; $T_0(\vec{v})$ to denote the set of $\vec{w} \in T - \{\vec{v}\}$ that is 0-common to \vec{v} . Then from the assumption that $m - n \geq m/c$, it follows that

$$\sum_{\vec{v} \in T} \sum_{\vec{w} \in T_2(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \leq \left\{ \frac{c^4(1-\delta)^2}{m^5} n^2 + \frac{2c^3(1-\delta)}{m^4} n \right\} |T|; \quad (7)$$

$$\sum_{\vec{v} \in T} \sum_{\vec{w} \in T_1(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \leq \left\{ \frac{4c^4\delta(1-\delta)^2}{m^6} n^3 + \frac{4c^3\delta(1-\delta)}{m^5} n^2 + \frac{4c^3\delta}{m^5} n \right\} |T|; \quad (8)$$

$$\sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \leq \mathbf{E}^2[Z] + \left\{ \frac{2c^4\delta^2(1-\delta)}{m^6} n^3 + \frac{c^4\delta^2}{m^6} n^2 \right\} |T|. \quad (9)$$

The proofs of Inequalities (7), (8), and (9) are shown in Subsections A.1, A.2, and A.3, respectively. Thus from Inequalities (5), (6), (7), (8), and (9), it follows that

$$\begin{aligned} \mathbf{Var}[Z] &\leq \mathbf{E}[Z] - \mathbf{E}^2[Z] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T - \{\vec{v}\}} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \\ &= \mathbf{E}[Z] - \mathbf{E}^2[Z] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_2(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_1(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \\ &\leq \frac{c^2|T|}{m^3} + \left\{ \frac{c^4(1-\delta)^2}{m^5} n^2 + \frac{2c^3(1-\delta)}{m^4} n \right\} |T| \end{aligned}$$

$$\begin{aligned}
& + \left\{ \frac{4c^4\delta(1-\delta)^2}{m^6}n^3 + \frac{4c^3\delta(1-\delta)}{m^5}n^2 + \frac{4c^3\delta}{m^5}n \right\} |T| \\
& + \left\{ \frac{2c^4\delta^2(1-\delta)}{m^6}n^3 + \frac{c^4\delta^2}{m^6}n^2 \right\} |T| \\
& \leq \frac{c^2|T|}{m^3} \left\{ 1 + \frac{c^2(1-\delta)^2}{m^2}n^2 + \frac{2c(1-\delta)}{m}n + \frac{4c^2\delta(1-\delta)^2}{m^3}n^3 \right. \\
& \quad \left. + \frac{4c\delta(1-\delta)}{m^2}n^2 + \frac{4c\delta}{m^2}n + \frac{2c^2\delta^2(1-\delta)}{m^3}n^3 + \frac{c^2\delta^2}{m^3}n^2 \right\} \\
& \leq \frac{c^2|T|}{m^3} \left\{ 1 + (c-1)^2(1-\delta)^2 + 2(c-1)(1-\delta) + \frac{4(c-1)^3\delta(1-\delta)^2}{c} \right. \\
& \quad \left. + \frac{4(c-1)^2\delta(1-\delta)}{c} + \frac{4(c-1)\delta}{m} + \frac{2(c-1)^3\delta^2(1-\delta)}{c} + \frac{(c-1)^2\delta^2}{m} \right\}, \tag{10}
\end{aligned}$$

where Inequality (10) follows from the assumption that $m-n \geq m/c$, i.e., $cn/m \leq c-1$. Thus it follows that $\mathbf{Var}[Z] \leq d|T|/m^3$ for some constant d that is determined by the constants $0 < \delta < 1$ and $c > 1$. Then from Inequalities (1), (2), and (4), we finally have that

$$\Pr[Z = 0] \leq \frac{\mathbf{Var}[Z]}{\mathbf{E}^2[Z]} \leq \frac{d|T|}{m^3} \cdot \frac{m^6}{|T|^2} = \frac{dm^3}{|T|} \leq \frac{3dm^3}{\delta^2(1-\delta)^2n^4} = O\left(\frac{m^3}{n^4}\right),$$

which implies that $\Pr[Z = 0] = o(1)$ for any $m \geq n$ with $m/n^{4/3} = o(1)$. Therefore, if $m/n^{4/3} = o(1)$, then with probability $1 - o(1)$, we have $Z > 0$, that is, $G = (V, E)$ contains G'_1 as a subgraph. \blacksquare

5 Upper Bounds for the 2-Weighted Popular Matching Problem

As shown in Theorem 4.1, a random instance of the complete and strict 2-weighted popular matching problem has a 2-weighted popular matching with probability $o(1)$ if $m/n^{4/3} = o(1)$. Here we consider roughly opposite case, i.e., $n^{4/3}/m = o(1)$, and prove that a random instance has a 2-weighted popular matching with probability $1 - o(1)$.

First we show the following lemma that will greatly simplify our later analysis.

Lemma 5.1 *If $n/m = o(1)$, then a random instance $G = (V, E)$ of the fs-relation graphs contains a cycle as a subgraph with probability $o(1)$.*

Proof: For each $\ell \geq 2$, let C_ℓ be a cycle with ℓ vertices and ℓ edges, and $\mathcal{E}_\ell^{\text{cyc}}$ be the event that a random fs-relation graph $G = (V, E)$ contains a cycle C_ℓ . Then from the assumption that $m - n \geq m/c$ for some constant $c > 1$, it follows that

$$\begin{aligned}
\Pr[G \text{ contains a cycle}] &= \Pr\left[\bigcup_{\ell \geq 2} \mathcal{E}_\ell^{\text{cyc}}\right] \leq \sum_{\ell \geq 2} \Pr[\mathcal{E}_\ell^{\text{cyc}}] \\
&\leq \sum_{\ell \geq 2} \left\{ \frac{1}{2^\ell} \ell! \binom{m}{\ell} \ell! \binom{n}{\ell} \left(\frac{1}{m-n}\right)^{2\ell} \right\} \leq \sum_{\ell \geq 2} \left\{ \frac{1}{2^\ell} m^\ell n^\ell \left(\frac{c}{m}\right)^{2\ell} \right\} \\
&= \sum_{\ell \geq 2} \frac{1}{2^\ell} \left(\frac{c^2 n}{m}\right)^\ell \leq \sum_{\ell \geq 2} \left(\frac{c^2 n}{m}\right)^\ell = \frac{c^4 n^2}{m^2} \sum_{h \geq 0} \left(\frac{c^2 n}{m}\right)^h = O\left(\frac{n^2}{m^2}\right),
\end{aligned}$$

where the last equality follows from the assumption that $n/m = o(1)$ and $c > 1$ is a constant. Thus it follows that if $n/m = o(1)$, then a random fs-relation graph $G = (V, E)$ contains a cycle as a subgraph with probability $o(1)$. \blacksquare

Theorem 5.1 *If $n^{4/3}/m = o(1)$, then a random instance of the complete and strict 2-weighted popular matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching with probability $1 - o(1)$.*

Proof: Consider a random fs-relation graph $G = (V, E)$ corresponding to a random instance of the complete and strict 2-weighted popular matching problem. By Lemma 5.1 and the assumption that $n^{4/3}/m = o(1)$, we know that the fs-relation graph $G = (V, E)$ contains bad subgraphs G_2 or G_3 of Figure 1 with vanishing probability $o(1)$. Thus in the rest of the proof, we estimate the probability that the graph $G = (V, E)$ contains a bad subgraph G_1 of Figure 1.

For any $\ell \geq 4$, let P_ℓ be a path with $\ell + 1$ vertices and ℓ edges, and $\mathcal{E}_\ell^{\text{path}}$ be the event that $G = (V, E)$ contains a path P_ℓ . It is obvious that a path P_ℓ is a bad subgraph G_1 for each $\ell \geq 4$. Then from the assumption that $m - n \geq m/c$ for some constant $c > 1$, it follows that

$$\begin{aligned} \Pr[G \text{ contains a bad subgraph } G_1] &= \Pr\left[\bigcup_{\ell \geq 4} \mathcal{E}_\ell^{\text{path}}\right] \leq \sum_{\ell \geq 4} \Pr[\mathcal{E}_\ell^{\text{path}}] \\ &\leq \sum_{\ell \geq 4} \left\{ \frac{1}{(m-n)^{2\ell}} (\ell+1)! \binom{m}{\ell+1} \ell! \binom{n}{\ell} \right\} \\ &\leq \sum_{\ell \geq 4} \left\{ \left(\frac{c}{m}\right)^{2\ell} m^{\ell+1} n^\ell \right\} = \frac{c^8 n^4}{m^3} \sum_{h \geq 0} \left(\frac{c^2 n}{m}\right)^h = O\left(\frac{n^4}{m^3}\right), \end{aligned}$$

where the last equality follows from the assumption that $n^{4/3}/m = o(1)$ and that $c > 1$ is a constant. Notice that $n/m = o(1)$ if $n^{4/3}/m = o(1)$. Thus from Lemma 5.1 and Corollary 3.1, it follows that if $n^{4/3}/m = o(1)$, then a random instance of the complete and strict 2-weighted popular matching problem has a 2-weighted popular matching with probability $1 - o(1)$. ■

6 Concluding Remarks

In this paper, we have analyzed the 2-weighted matching problem, and have shown that (Theorem 4.1) if $m/n^{4/3} = o(1)$, then a random instance of the complete and strict 2-weighted popular matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching with probability $o(1)$; (Theorem 5.1) if $n^{4/3}/m = o(1)$, then a random instance of the complete and strict 2-weighted popular matching problem with $w_1 \geq 2w_2$ has a 2-weighted popular matching with probability $1 - o(1)$. These results imply that there exists a threshold $m \approx n^{4/3}$ to admit 2-weighted popular matchings, which is quite different from the case for the single category popular matching problem due to Mahdian [7].

Theorem 4.1 can be trivially generalized to any multiple category case; that is, with the same proof, we have the following bound.

Theorem 6.1 *For any integer $k > 2$, if $m/n^{4/3} = o(1)$, then a random instance of the complete and strict k -weighted popular matching problem with $w_i \geq 2w_{i+1}$ ($1 \leq i \leq k-1$) has a k -weighted popular matching with probability $o(1)$.*

Then an interesting problem is to show some upper bound result by generalizing Theorem 5.1 for any integer $k > 2$, maybe under the condition that $w_i \geq 2w_{i+1}$ for all i , $1 \leq i \leq k-1$.

References

- [1] D.J. Abraham, N. Chen, V. Kumar, and V. Mirrokni. Assignment Problems in Rental Markets. In *Proc. Internet and Network Economics*, Lecture Notes in Computer Science 4286, pp.198-213, 2006.

- [2] D.J. Abraham, R.W. Irving, T. Kavitha, and K. Mehlhorn. Popular Matchings. In *Proc. of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp.424-432, 2005.
- [3] D.J. Abraham and T. Kavitha. Dynamic Matching Markets and Voting Paths. In *Proc. of the 10th Scandinavian Workshop on Algorithm Theory*, Lecture Notes in Computer Science 4059, pp.65-76, 2006.
- [4] N. Alon and J. Spencer. *The Probabilistic Method*. John Wiley & Sons, 2000.
- [5] B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.
- [6] P. Gardenfors. Match Making: Assignment Based on Bilateral Preferences. *Behaviourial Sciences*, 20:166-173, 1975.
- [7] M. Mahdian. Random Popular Matchings. In *Proc. of the 7th ACM Conference on Electronic Commerce*, pp.238-242, 2006.
- [8] J. Mestre. Weighted Popular Matchings. In *Proc. of the 33rd International Colloquium on Automata, Languages, and Programming (Part I)*, Lecture Notes in Computer Science 4051, pp.715-726, 2006.
- [9] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

A Proofs of Inequalities

A.1 Proof of Inequality (7)

Let $\vec{v} = (x_1, x_2, y_1, y_2) \in T$. For each $\vec{w} = (x'_1, x'_2, y'_1, y'_2) \in T_2(\vec{v})$, let us consider the following cases: (case-0) $|\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 0$; (case-1) $|\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 1$. Let

$$\begin{aligned} T_2^0(\vec{v}) &= \{\vec{w} \in T_2(\vec{v}) : |\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 0\}; \\ T_2^1(\vec{v}) &= \{\vec{w} \in T_2(\vec{v}) : |\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 1\}. \end{aligned}$$

For each $\vec{v} \in T$, it is immediate to see that $T_2^0(\vec{v}), T_2^1(\vec{v})$ is the partition of $T_2(\vec{v})$, and from the definitions of $T_2^0(\vec{v})$ and $T_2^1(\vec{v})$, we have that $|T_2^0(\vec{v})| \leq n_2^2$; $|T_2^1(\vec{v})| \leq 2n_2$. So from the assumption that $m - n \geq m/c$ for some constant $c > 1$, it follows that for each $\vec{v} \in T$,

$$\begin{aligned} \sum_{\vec{w} \in T_2^0(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &\leq \sum_{\vec{w} \in T_2^0(\vec{v})} \frac{1}{m} \left(\frac{1}{m - n_1} \right)^4 \leq \sum_{\vec{w} \in T_2^0(\vec{v})} \frac{1}{m} \left(\frac{1}{m - n} \right)^4 \\ &\leq \sum_{\vec{w} \in T_2^0(\vec{v})} \frac{1}{m} \left(\frac{c}{m} \right)^4 = \frac{c^4}{m^5} |T_2^0(\vec{v})| \leq \frac{c^4}{m^5} n_2^2 \\ &= \frac{c^4(1 - \delta)^2}{m^5} n^2; \end{aligned} \tag{11}$$

$$\begin{aligned} \sum_{\vec{w} \in T_2^1(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &\leq \sum_{\vec{w} \in T_2^1(\vec{v})} \frac{1}{m} \left(\frac{1}{m - n_1} \right)^3 \leq \sum_{\vec{w} \in T_2^1(\vec{v})} \frac{1}{m} \left(\frac{1}{m - n} \right)^3 \\ &\leq \sum_{\vec{w} \in T_2^1(\vec{v})} \frac{1}{m} \left(\frac{c}{m} \right)^3 = \frac{c^3}{m^4} |T_2^1(\vec{v})| \leq \frac{2c^3}{m^4} n_2 \\ &= \frac{2c^3(1 - \delta)}{m^4} n. \end{aligned} \tag{12}$$

Thus from Inequalities (11) and (12), we finally have that

$$\begin{aligned} \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_2(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &= \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_2^0(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_2^1(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \\ &\leq \sum_{\vec{v} \in T} \left\{ \frac{c^4(1 - \delta)^2}{m^5} n^2 + \frac{2c^3(1 - \delta)}{m^4} n \right\} \\ &= \left\{ \frac{c^4(1 - \delta)^2}{m^5} n^2 + \frac{2c^3(1 - \delta)}{m^4} n \right\} |T|. \end{aligned}$$

A.2 Proof of Inequality (8)

Let $\vec{v} = (x_1, x_2, y_1, y_2) \in T$. For each $\vec{w} = (x'_1, x'_2, y'_1, y'_2) \in T_1(\vec{v})$, we have the following cases: (case-0) $|\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 0$; (case-1) $|\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 1$; (case-2) $|\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 2$. Let

$$\begin{aligned} T_1^0(\vec{v}) &= \{\vec{w} \in T_1(\vec{v}) : |\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 0\}; \\ T_1^1(\vec{v}) &= \{\vec{w} \in T_1(\vec{v}) : |\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 1\}; \\ T_1^2(\vec{v}) &= \{\vec{w} \in T_1(\vec{v}) : |\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 2\}. \end{aligned}$$

For each $\vec{v} \in T$, it is immediate that $T_1^0(\vec{v}), T_1^1(\vec{v}), T_1^2(\vec{v})$ is the partition of $T_1(\vec{v})$, and from the definitions of $T_1^0(\vec{v})$, $T_1^1(\vec{v})$, and $T_1^2(\vec{v})$, we have that $|T_1^0(\vec{v})| \leq 4n_1 n_2^2$; $|T_1^1(\vec{v})| \leq 4n_1 n_2$; $|T_1^2(\vec{v})| \leq 4n_1$. So from the

assumption that $m - n \geq m/c$ for some constant $c > 1$, it follows that for each $\vec{v} \in T$,

$$\begin{aligned}
\sum_{\vec{w} \in T_1^0(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &\leq \sum_{\vec{w} \in T_1^0(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n_1} \right)^4 \leq \sum_{\vec{w} \in T_1^0(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n} \right)^4 \\
&\leq \sum_{\vec{w} \in T_1^0(\vec{v})} \frac{1}{m^2} \left(\frac{c}{m} \right)^4 = \frac{c^4}{m^6} |T_1^0(\vec{v})| \leq \frac{4c^4}{m^6} n_1 n_2^2 \\
&= \frac{4c^4 \delta (1 - \delta)^2}{m^6} n^3;
\end{aligned} \tag{13}$$

$$\begin{aligned}
\sum_{\vec{w} \in T_1^1(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &\leq \sum_{\vec{w} \in T_1^1(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n_1} \right)^3 \leq \sum_{\vec{w} \in T_1^1(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n} \right)^3 \\
&\leq \sum_{\vec{w} \in T_1^1(\vec{v})} \frac{1}{m^2} \left(\frac{c}{m} \right)^3 = \frac{c^3}{m^5} |T_1^1(\vec{v})| \leq \frac{4c^3}{m^5} n_1 n_2 \\
&= \frac{4c^3 \delta (1 - \delta)}{m^5} n^2;
\end{aligned} \tag{14}$$

$$\begin{aligned}
\sum_{\vec{w} \in T_1^2(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &\leq \sum_{\vec{w} \in T_1^2(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n_1} \right)^3 \leq \sum_{\vec{w} \in T_1^2(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n} \right)^3 \\
&\leq \sum_{\vec{w} \in T_1^2(\vec{v})} \frac{1}{m^2} \left(\frac{c}{m} \right)^3 = \frac{c^3}{m^5} |T_1^2(\vec{v})| \leq \frac{4c^3}{m^5} n_1 \\
&= \frac{4c^3 \delta}{m^5} n.
\end{aligned} \tag{15}$$

Thus from Inequalities (13), (14), and (15), we finally have that

$$\begin{aligned}
\sum_{\vec{v} \in T} \sum_{\vec{w} \in T_1(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &= \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_1^0(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_1^1(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_1^2(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \\
&\leq \sum_{\vec{v} \in T} \left\{ \frac{4c^4 \delta (1 - \delta)^2}{m^6} n^3 + \frac{4c^3 \delta (1 - \delta)}{m^5} n^2 + \frac{4c^3 \delta}{m^5} n \right\} \\
&= \left\{ \frac{4c^4 \delta (1 - \delta)^2}{m^6} n^3 + \frac{4c^3 \delta (1 - \delta)}{m^5} n^2 + \frac{4c^3 \delta}{m^5} n \right\} |T|.
\end{aligned}$$

A.3 Proof of Inequality (9)

Let $\vec{v} = (x_1, x_2, y_1, y_2) \in T$. For each $\vec{w} = (x'_1, x'_2, y'_1, y'_2) \in T_0(\vec{v})$, we have the following cases: (case-0) $|\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 0$; (case-1) $|\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 1$; (case-2) $|\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 2$. Let

$$\begin{aligned}
T_0^0(\vec{v}) &= \{\vec{w} \in T_0(\vec{v}) : |\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 0\}; \\
T_0^1(\vec{v}) &= \{\vec{w} \in T_0(\vec{v}) : |\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 1\}; \\
T_0^2(\vec{v}) &= \{\vec{w} \in T_0(\vec{v}) : |\{y_1, y_2\} \cap \{y'_1, y'_2\}| = 2\}.
\end{aligned}$$

For each $\vec{v} \in T$, it is immediate that $T_0^0(\vec{v}), T_0^1(\vec{v}), T_0^2(\vec{v})$ is the partition of $T_0(\vec{v})$. For any $\vec{w} \in T_0^0(\vec{v})$, it is obvious that $\Pr[Z_{\vec{v}} = 1 \wedge Z_{\vec{w}} = 1] = \Pr[Z_{\vec{v}} = 1] \times \Pr[Z_{\vec{w}} = 1]$, which implies that

$$\sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0^0(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] = \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0^0(\vec{v})} \Pr[Z_{\vec{v}} = 1 \wedge Z_{\vec{w}} = 1]$$

$$\begin{aligned}
&= \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0^0(\vec{v})} \Pr[Z_{\vec{v}} = 1] \times \Pr[Z_{\vec{w}} = 1] \\
&= \sum_{\vec{v} \in T} \Pr[Z_{\vec{v}} = 1] \sum_{\vec{w} \in T_0^0(\vec{v})} \Pr[Z_{\vec{w}} = 1] \\
&\leq \sum_{\vec{v} \in T} \Pr[Z_{\vec{v}} = 1] \sum_{\vec{w} \in T} \Pr[Z_{\vec{w}} = 1] \\
&= \mathbf{E}^2[Z].
\end{aligned} \tag{16}$$

From the definitions of $T_0^1(\vec{v})$ and $T_0^2(\vec{v})$, we have that $|T_0^1(\vec{v})| \leq 2n_1^2 n_2$; $|T_0^2(\vec{v})| \leq n_1^2$. Then from the assumption that $m - n \geq m/c$ for some constant $c > 1$, it follows that for each $\vec{v} \in T$,

$$\begin{aligned}
\sum_{\vec{w} \in T_0^1(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &= \sum_{\vec{w} \in T_0^1(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n_1} \right)^4 \leq \sum_{\vec{w} \in T_0^1(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n} \right)^4 \\
&\leq \sum_{\vec{w} \in T_0^1(\vec{v})} \frac{1}{m^2} \left(\frac{c}{m} \right)^4 = \frac{c^4}{m^6} |T_0^1(\vec{v})| \leq \frac{2c^4}{m^6} n_1^2 n_2 \\
&= \frac{2c^4 \delta^2 (1 - \delta)}{m^6} n^3,
\end{aligned} \tag{17}$$

$$\begin{aligned}
\sum_{\vec{w} \in T_0^2(\vec{v})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &= \sum_{\vec{w} \in T_0^2(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n_1} \right)^4 \leq \sum_{\vec{w} \in T_0^2(\vec{v})} \frac{1}{m^2} \left(\frac{1}{m - n} \right)^4 \\
&\leq \sum_{\vec{w} \in T_0^2(\vec{v})} \frac{1}{m^2} \left(\frac{c}{m} \right)^4 = \frac{c^4}{m^6} |T_0^2(\vec{v})| \leq \frac{c^4}{m^6} n_1^2 \\
&= \frac{c^4 \delta^2}{m^6} n^2.
\end{aligned} \tag{18}$$

Thus from Inequalities (16), (17), and (18), we finally have that

$$\begin{aligned}
\sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0(\vec{w})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] &= \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0^0(\vec{w})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0^1(\vec{w})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] + \sum_{\vec{v} \in T} \sum_{\vec{w} \in T_0^2(\vec{w})} \mathbf{E}[Z_{\vec{v}} Z_{\vec{w}}] \\
&\leq \mathbf{E}^2[Z] + \sum_{\vec{v} \in T} \left\{ \frac{2c^4 \delta^2 (1 - \delta)}{m^6} n^3 + \frac{c^4 \delta^2}{m^6} n^2 \right\} \\
&= \mathbf{E}^2[Z] + \left\{ \frac{2c^4 \delta^2 (1 - \delta)}{m^6} n^3 + \frac{c^4 \delta^2}{m^6} n^2 \right\} |T|.
\end{aligned}$$